

The Likelihood Principle: Objectivity and the Values and Science Debate

Sean Boivin

A Major Research Paper

in

the Department of Philosophy

Under the supervision of

Dr. Matthew J. Barker

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Arts (Philosophy) at

Concordia University

Montreal, Quebec, Canada

February 2018

© Sean Boivin

ABSTRACT

The Likelihood Principle: Objectivity and the Values and Science Debate

Sean Boivin

This paper focuses on the debate of underdetermination in science, and asks the descriptive question: is objectivity possible in science? I introduce the problem of underdetermination in science and articulate a related argument presented by philosopher Helen Longino against the possibility for objectivity (traditionally understood) in science. In opposition to Longino, I aim to salvage the possibility of important objectivity. I begin from Likelihoodism – a normative view about the form that evidential reasoning should take. After presenting different defenses of that view, I show how it implies a descriptive claim – the Likelihood Principle – that opposes Longino’s cynicism about the descriptive possibility of objectivity in science. The Likelihood Principle compares the likelihoods of two hypotheses in relation to a body of evidence and says which hypothesis (if any) is consequently favored. I argue that “favours” be interpreted as “objectively favours”, implying it is possible for some evidence to objectively favour one hypothesis over another without appeal to values. In addition to arguing that we should then infer a descriptive objectivism from this, I interpret a case-study using the Likelihood Principle to illustrate how applications of it can be objective. I discuss what follows for the debate in the values and science literature, including what follows with respect to Longino’s views.

ACKNOWLEDGEMENTS

I express my sincere gratitude to my supervisor, Dr. Matthew Barker for his guidance and mentorship over the course of my graduate studies. His support and feedback have been invaluable in my research.

I would also like to thank Dr. Ingo Brigandt for his hospitality during my research stay in Edmonton, Alberta, and for his insightful comments that have helped guide my research and my writing of this MRP.

Finally, I am grateful for the love I receive from my parents, siblings, and my partner. Their continued support and encouragement has carried me throughout this journey.

Table of Contents

Section 1.1 - Introduction	1
Section 1.2 - Longino and Underdetermination	3
Section 2.1: Likelihoodism	8
Section 2.2: The Likelihood Principle.....	11
Section 2.3: Likelihoodism as Justified.....	11
Section 2.3.1: Inductive Risk and Royall’s Three Questions.....	20
Section 3.1: A Case Study	22
Section 3.2: Applying the Likelihood Principle.....	22
Section 3.3: Objective Favouring.....	26
Section 4.1: Objection.....	27
Section 4.2: Response	32
Conclusion	33
Endnotes.....	34
Bibliography	45

Section 1.1 - Introduction

A traditional view about values in science says that “non-epistemic” values¹ (e.g., ethical, political, or religious concerns) can and should play roles only at the periphery of science, outside the “context of justification” in which evidence can and should objectively and epistemically relate to hypotheses and our beliefs about them.² In recent years authors have challenged this, arguing that moral, social and political values *should* play a role in the context of justification.³ Others have argued that non-epistemic values *necessarily* influence scientific reasoning, and thus justification.⁴ The more traditional and opposing view that non-epistemic values needn’t and shouldn’t have any relevance for assessing the degree to which scientific hypotheses are confirmed is dubbed the *value-free ideal* for science.

It is now widely accepted that “epistemic” values (e.g., predictive accuracy, explanatory power, and consistency) have a legitimate role to play throughout all aspects of scientific reasoning.⁵ Moreover, most scholars agree that “non-epistemic” values can appropriately influence many aspects of science, including choices about what projects to pursue, what ethical constraints to place on scientific methodologies⁶ and what counts as a reliable finding.⁷ Current debate among philosophers of science revolves primarily around the question of whether non-epistemic values also have a legitimate role to play at the center of *scientific reasoning*.⁸ Numerous strategies for answering this question have recently been proposed.⁹ One is to rely strictly on a social account of the nature of science¹⁰ and argue that non-epistemic factors have a legitimate and unavoidable role to play in evaluating scientific claims because scientific claims are underdetermined by purely epistemic considerations.¹¹ For example, in her book *Science as Social Knowledge*, philosopher Helen Longino rejects the value-free ideal for science and argues

that data stand in evidential relationships to theories or hypotheses only in the context of auxiliary assumptions that should be collectively assessed against both epistemic and non-epistemic values.

I believe that some challenges to the value-free ideal are well-taken and important but also that some modest but important objectivity can and should be retained. To more exactly spell this out, my paper will first focus on a descriptive question: is objectivity possible in science? I next introduce the related issue of underdetermination in science, then present Longino's arguments against the possibility for objectivity (traditionally understood) in science (Section 1.2). In opposition to Longino on my descriptive question I argue for an objectivist form of Likelihoodism (Section 2). Although this is a normative view about the form that evidential reasoning should take, it also implies a descriptive claim – the Likelihood Principle – that opposes Longino's cynicism about the descriptive possibility of objectivity in science. To see this, consider what Likelihoodism says:

If a rational agent is determining what evidential relationship holds between an observation, O , a hypothesis, H_1 , and another hypothesis, H_2 , then she should do so in accord with the Likelihood Principle: O favours H_1 over H_2 *iff* $P(O|H_1) > P(O|H_2)$.

This clarifies that Likelihoodism, as a normative view about how one *should* reason in particular cases, makes use of the Likelihood Principle, a descriptive claim that says exactly when O *does and does not* favour H_1 over H_2 .¹² I will show that this descriptive principle disagrees with Longino's cynicism because we should interpret the favouring it describes as objective. And by implying that O objectively favours the hypothesis on which O was more probable, the principle implies that such objectivity is, as a matter of descriptive fact, possible.

While arguing in this way, I will be presupposing Likelihoodism is justified and present different strategies authors invoke in its defense. In addition to arguing that we should then infer a descriptive objectivism from this, I interpret a case-study using the Likelihood Principle (Section 4), in order to illustrate how applications of it can be objective. I discuss what follows for the debate in the values and science literature, including what follows with respect to Longino's views. Authors have not appreciated the problems that Likelihoodism poses for criticisms of the value-free ideal.

Section 1.2 - Longino and Underdetermination

The descriptive question about whether objectivity is possible in science is often addressed within debate about so-called underdetermination, which concerns the role of auxiliary assumptions in scientific reasoning about hypotheses and evidence. It is now widely appreciated that for nearly any particular scientific hypothesis, observations alone are not enough to determine whether the hypothesis is supported.¹³ Observations themselves leave the issue of the hypothesis' support underdetermined. This is because it is very rare for a hypothesis to *all by itself* make predictions about which observations will be obtained. To make predictions, hypotheses nearly always must be conjoined with auxiliary assumptions. For instance, the hypothesis that stormy weather is caused by falling air pressure does not, by itself, predict that a barometer reading will be observed to fall when a storm rolls in. To make that prediction, we must add the following sort of auxiliary assumptions to the hypothesis: the barometer is working correctly, observations of the barometer are accurate, the barometer and the storm are within the same weather system, etc. When a prediction fails, this suggests a problem somewhere in the total package of hypothesis + auxiliary assumptions, but the failure alone is silent about where in the package the problem resides. Suppose we consult a barometer as a storm rolls in, and observe

that it is *not* falling. The hypothesis about stormy weather plus usual auxiliary assumptions fail to predict this. But whether this evidentially counts against the hypothesis in particular, or against it plus some of the auxiliary assumptions, or against just some of those assumptions, is left so far undetermined. Perhaps the barometer is in a climate controlled room, in a house in the storm, such that the problem resides solely with the assumption that the barometer and the storm are within the same weather system.

It is common for authors to move from those widely accepted points about underdetermination, to the following line of argument. If some observation objectively favors a particular hypothesis, then it is possible to isolate the hypothesis and each of the associated auxiliary assumptions and exhaustively test or otherwise justify each of these individually. But it is not possible to do that. Therefore, an observation cannot objectively favour a hypothesis.¹⁴

This argument is supposed to apply to all sciences but is most prevalently discussed in connection with the human and social sciences. Helen Longino leverages it in her 1990 book, while referring to auxiliary assumptions as background beliefs. She argues that if background beliefs are necessary to bridge the gap between evidence and hypotheses, so that what in fact counts as evidence is relativized to background beliefs, then hypothesis *acceptance* on the basis of evidence is also in fact relativized.¹⁵ She realizes that an infinite regress then threatens: to strengthen evidence so that it justifies accepting a hypothesis, we will often need to test the background beliefs with respect to which existing evidence is relativized; but to do that for the background beliefs, we will have to rely on and test still other background beliefs, and so on.

But Longino claims science can and often does effectively address this threat of regress via *contextual empiricism*. This view notes that science is conducted within a *social* context, wherein background assumptions are articulated, accessed, and scrutinized from distinct and

complimentary points of view by a *diverse community* of scientists, philosophers and others. Mechanisms of peer review allow for this, for example. Longino posits that such intersubjective mechanisms are key for the rationality of science, rather than subject-independent objectivity. Such objectivity was traditionally thought furnished by having evidence alone, independent of judgments by subjects, determine the support or lack thereof for hypotheses. Longino believes the underdetermination problem shows that objectivity in that traditional sense is impossible. Not only does she then advise that we then turn instead to intersubjective mechanisms to help secure rationality in science, but also she implies that we should change what we mean by 'objective' so that it no longer refers to the traditional subject-independent concept and instead refers to the intersubjective concept.¹⁶ She believes that diversity in values held by different subjects of scientific communities is important for achieving increasing degrees of objectivity (understood in her new sense) via the social mechanisms she clarifies.

Longino further develops a conception of scientific *practice* in her 2002 book *The Fate of Knowledge*, claiming that, traditionally understood, science aims to accept true statements, and often succeeds in this. But philosophers have typically thought that scientists do and should attempt to achieve this aim individualistically, with acceptance by the scientific community being acceptance by the individuals who comprise that community. Her alternative conclusion draws from her 1990 book: acceptance of scientific statements is instead the result of *important interactions* between subjects within the scientific community, through debate, collaboration, instruction, and so forth.

Accordingly, Longino proposes to amend the notion of rules of method in science traditionally understood. She thinks the traditional rules of method were originally adopted on the basis of the belief that following them reliably generates and sustains true *beliefs*. To this,

Longino adds that there are rules governing the *interactions* among scientists as well, and she proposes four ideal norms that, when taken together, she thinks address the problem of underdetermination: public venues for criticism; uptake of criticism; publicly recognized standards for evaluating theories, hypotheses and observation practices; and tempered equality across class, genders and communities.¹⁷ Background beliefs surviving critical scrutiny under these conditions satisfy the notion Longino calls “epistemic acceptability”: a combination of the empiricist notion of justification by empirical data and the social norms applying to interaction within communities that constitutes reasoning.¹⁸ Longino thinks that operating in accord with the four ideal norms she presents is what it means to be objective (in her sense) and thinks of these norms as hypotheticals, rather than categorical.¹⁹ Longino labels her view “sociopragmatism”, a parent view that includes but also expands upon her view of contextual empirism I introduced above. Per this parent view, knowledge is partial and reflects a point of view, is plural with different practices producing different sorts of knowledge about the same phenomenon, is provisional,²⁰ and does not assume either monism or pluralism but is open to the possibility of even a strong form of pluralism in which different background assumptions could lead to *incompatible* and *irreconcilable* versions of the same phenomenon.

More recently in her 2013 book, while focused on examples from contemporary research programs studying human aggression and sexuality, she argues for a strong pluralism in which background assumptions lead researchers to incompatible and irreconcilable versions of the same phenomenon. Devoting one chapter to each of the different research programs on human sexuality and aggression that she considers, Longino cites the concepts, methodologies and assumptions employed by each program as well as the criticisms raised against conclusions of researchers.

A main claim in her argument is that the theories embedded within a research program are irreconcilable and empirically incommensurable with theories embedded in other research programs.²¹ She argues this is because even while different research programs may be attempting to answer the same general question (i.e., What is the cause of human aggression and sexuality?), theories embedded within each program rely on different background assumptions, employ different evidence-gathering procedures and data-interpretation techniques, use different theoretical concepts than theories embedded within other research programs, and possess different research aims. For example, quantitative behavioral genetics (QBG) focuses on the whole genome, as well as shared and nonshared environmental factors, while drawing upon twin studies and adoption studies (MZ/DZ comparisons) to infer heritability estimates for a given trait. Developmental systems theorists, in contrast, treat prior states of entire developmental systems as causes of changes of state in those systems, and they rely mostly on animal subjects for testing while focusing primarily on species specific traits.²² Social-environmental approaches are a third type and focus on more fine-grained environmental factors (within what QBG calls “non-shared environment”), often with the aim of determining the effectiveness of social interventions rather than the underlying causes and with the primary research aim of justifying social work and clinical psychology.

Due to this heterogeneity of research programs (the use of different background assumptions, methodologies, research aims, experimental designs, data interpretation techniques and so on) many research programs are, according to Longino, incompatible with one another and cannot be simply integrated into a single fundamental account. More importantly, such incompatibilities reveal how different and incompatible standards for evaluating evidence are in use across research programs, rendering them empirically incommensurable. An important

implication Longino draws from this is that scrutinizing evidence can only occur from *within* the research program that produced it, denying the possibility for cross-approach empirical evaluation.²³ The resulting picture Longino paints for her readers is that each research program provides only partial knowledge from the purview of a particular set of parts that together comprise a research program.²⁴ More importantly, “better” knowledge is produced by “[r]efining and improving methods” *within* a particular program, but any disagreement or contradiction *across* research programs cannot be settled by evidence.²⁵

To the extent that she thinks high degrees of objectivity are achievable only when values play roles in the research questions being asked, and in the kinds of experiments being conducted and data sets collected, she gives an answer to my descriptive question of whether objectivity is possible in science – Longino believes that in the vast majority of cases, it is not possible for evidence to objectively favor a hypothesis, unless by the term “objectively” we include appeal to a diverse set of values that interact as checks and balances. To respond to Longino’s challenges, in the next section I discuss the Likelihoodist alternative to Longino’s views, an alternative that preserves a role for a more traditional concept of objectivity in science.

Section 2.1: Likelihoodism

This section discusses the reasons for believing Likelihoodism is true and that its prescription can, when followed, yield support for hypotheses that is objective. I argue that the reasons for believing Likelihoodism is true *imply* that we also have reason to believe some applications of the Likelihood Principle reveal *objective* favouring relations.²⁶ This is because of how Likelihoodism makes use of the Likelihood Principle, and the nature of that principle.

Where O is some data, H1 is one hypothesis, and H2 is another hypothesis, the Likelihood Principle says:

O favors H1 over H2 iff: $\Pr(O|H1) > \Pr(O|H2)$ ²⁷

In part, the Likelihood Principle says that if an event is more probable under H1 than H2, then occurrence of that event is evidence supporting H1 over H2. An application of the Principle compares the Likelihoods of two hypotheses in relation to data and says which hypothesis (if any) is consequently *favored* by those data.²⁸ Take for example the following two hypotheses: “it will rain tomorrow in Montreal” and “it will be clear and sunny all day tomorrow in Montreal”. Assign “H1” as the former hypothesis and “H2” as the latter. Now say that at ten o’clock today, reliably sourced meteorological data and models give a forecast that there is an eighty-percent chance of rain tomorrow and a twenty-percent chance of no rain tomorrow. Let “O” be the weather forecast. In such a case, applying the Likelihood Principle says that O favours H1 over H2, since H1 predicts that a reliable forecast will call for a strong chance of rain. The Likelihood Principle gives exact expression to our powerful intuition that says an observation is evidence supporting the hypothesis that, of those compared, predicted the observation more strongly. On such grounds the principle has been defended at length as a general tool for both formal and informal reasoning about hypothesis ranking.²⁹

The concept of objectivity that interests me is epistemic. It is about whether data, such as O, support the truth of some hypothesis over another in an objective sense. Is such objectivity of support possible in science? Let's assume O is some collected data that accurately represents the world as intended. Also assume that O is in fact more probable according to H1 than to H2, and so the Likelihood Principle implies that O favours H1 over H2. Now, what further condition(s) suffice(s) along with those assumptions for the favouring relation to be objective – for the favouring to amount to H1 being objectively supported over H2 by O? Here is a further condition that I propose to be sufficient (given the other assumptions) for such objectivity of favouring: O

would favour H1 over H2 even if all parties involved *denied* this after accepting that O is accurate as intended and is more probable on H1 than H2.

The forecast for rain example can illustrate this condition. Assume that O, the forecast expecting rain, accurately represents the predictions to come from the weather data and models. And because those data and models are reliable, their indication of rain is more probable on the hypothesis that it will rain tomorrow (H1) than on the hypothesis that it will be clear and sunny all day (H2). So the Likelihood Principle implies that O favours H1 over H2. Now imagine that everybody who learns of the forecast denies that it favours rain tomorrow. They accept that the forecast accurately reflects the data and models, and that the forecast is more probable on the hypothesis of rain tomorrow. But they uniformly deny that this is enough for the forecast to be evidence that favours the truth of the hypothesis of rain tomorrow. On my proposals, if the forecast nonetheless would still favour the hypothesis of rain, despite the uniform denial of this, then this favouring is objective: O is objectively supporting the truth of H1 over the truth of H2.

In contrast, if instead there would be no such favouring upon people uniformly denying such favouring, it would seem the favouring wasn't objective after all. But I will remain agnostic on this point. I am proposing my condition as sufficient in such cases for objectivity, while setting aside whether it is also necessary.

The criterion of (sufficient condition for) objectivity I have formulated and will defend here is important because much of the literature of values and science focuses on novel concepts of objectivity that do away with many of the traditional distinctions between the discovery, justification, and application aspects of science. The criterion I propose is also one that authors like Longino imply is impossible to satisfy. And so if successful, I will have defended modest

but important objectivity that has been overlooked or discarded too quickly within the science and values literature.

Section 2.2: The Likelihood Principle

One way of satisfying the criterion of objectivity I proposed is by applying the Likelihood Principle, which is a descriptive component of Likelihoodism. As will clarify, Likelihoodism also contains a normative component, saying that a person who is reasoning evidentially *should* reason in accord with the Likelihood Principle when the conditions for that principle's application hold. By way of actual examples and thought experiments, I will show how this view fits our least controversial judgments in certain cases better than rival approaches. I summarize how other authors implement this strategy when arguing for the truth of Likelihoodism, and I judge that they have thereby provided adequate defense of Likelihoodism over rival approaches. With our reasons for believing Likelihoodism thus clarified, my key claim is that we *thereby* also have reason to believe (even if we did not previously appreciate this) that applications of the Likelihood Principle can reveal favouring relations that are objective in the sense captured by my criterion of objectivity. This is because of how the Likelihood Principle is a component of Likelihoodism.

Section 2.3: Likelihoodism as Justified

So what are the reasons to believe Likelihoodism?

The most decisive type of defense that authors have provided for the Likelihood Principle is a mathematical proof. But this is a very particular goal-relative proof. The proof shows that if your goal is to determine which of two models is more predicatively accurate in a specified technical sense, then applying the Likelihood Principle is *the* way to achieve this goal.³⁰

When turning to a more general defense of the Likelihoodism that is not restricted to the goal of achieving a technical sort of predictive accuracy, two general strategies are available. One strategy is by way of actual examples and thought experiments. This involves arguing that Likelihoodism best systematizes judged examples. It compares our judgements about the evidence within a range of cases and shows how Likelihoodism does the best job at capturing our judgements in those cases.

The other general strategy for defending Likelihoodism resembles strategies in meta-ethics: you ask what is it that *makes* Likelihoodism true. The idea is to show that Likelihoodism is made true independently of our proclivities, by some real feature of the world that includes the Likelihood Principle. This strategy – a kind of *meta-epistemics* – is not one I employ. For the purposes of my paper, I adopt the former strategy.

There are two main alternatives to Likelihoodism. One is Bayesianism, which is compatible with some but not all parts of Likelihoodism, and the other is Frequentism such as significance testing, which is quite different from Likelihoodism. These are the two main alternatives to check Likelihoodism against, when claiming it is the best of available views³¹.

The Bayesianism approach involves interpreting how data bear on hypotheses. Central to Bayesianism is Bayes' theorem, a mathematical theorem derived from the axioms of Probability theory. Where O is some data, H1 is one hypothesis and (H | O) is the probability of a hypothesis given some observation, Bayes Theorem says:

$$\Pr(H | O) = [\Pr(H) \Pr(O | H)] / \Pr(O)$$

Bayes' theorem derives a posterior probability by multiplying the prior probability of H (i.e., Pr(H)) by the probability of the observation given the hypothesis (i.e., Pr(O|H)), then dividing by the prior probability of the observation (i.e., Pr(O)). Bayesianism is an epistemology that adds to

the merely mathematical Bayes' theorem. Bayesianism proposes that the posterior probability be understood as what your updated degree of belief in the hypothesis *should* be as you acquire new evidence. Bayesianism says:

O confirms H if and only if $\Pr(H | O) > \Pr(H)$

O disconfirms H if and only if $\Pr(H | O) < \Pr(H)$

O is conformationally irrelevant to H iff $\Pr(H | O) = \Pr(H)$

Bayesian confirmation is probabilistic, not dichotomous; it tells you how probably true or false a belief is, rather than simply asserting that it is true or that it is false. And as new observations come in, it tells you how these probabilities change, how you should change your *degrees of belief*. Confirmation (and disconfirmation) then involves *comparing* new and old degrees of belief.

Those who employ Bayesianism think it is an objective way of updating the probability of or degree of belief in a theory when certain conditions are met³². For example, imagine you're a doctor determining whether your patient has tuberculosis. Prior to administering a diagnostic test, you examine the patient and refer to frequency data concerning the prevalence of tuberculosis in the population to which the patient belongs. Say the disease is very rare, 1 in 1000, and the chance of error for the diagnostic test as either a false positive or false negative is extremely low, .01. Given these values, the posterior probabilities of $\Pr(\text{tuberculosis} | + \text{ result})$ and $\Pr(\text{no tuberculosis} | - \text{ result})$ are both large. The objectivity of the quantities that figure into the question of whether your patient has tuberculosis is apparent. When we say:

$\Pr(\text{no-tuberculosis} | - \text{ result}) > \Pr(\text{tuberculosis} | - \text{ result}),$

it is not merely that we *hope* that the one probability is of greater value than the other one, but

that the one probability *really is* greater than the other. As Sober states, “the objective component is substantial and compelling”.³³

Bayesianism can be a useful way of justifying the degrees of belief we should have in scientific theories or results of reliable³⁴ diagnostic tests. But there are many cases where applying Bayes’ theorem is problematic. When the data for estimating values of prior probabilities come from small or unrepresentative samples, scientists lack ground for reaching a justified consensus on what values to assign to the priors using the data alone. In these cases, some become critical of Bayesianism and its reliance on assigning values to priors. But there is a fallback position. It contains only some of the strengths of Bayesianism, to which one can retreat when the weaknesses of assigning priors make it wise to abandon Bayesianism. That position is *Likelihoodism*.

To briefly recall some of Section 1, Likelihoodism proposes the use of the Likelihood Principle and is a normative epistemic claim about how a person should reason evidentially. It says:

If a rational agent is determining what evidential relationship holds between O, H1 and H2, then she should do so in accord with the Likelihood Principle, i.e., she should believe that O favors H1 over H2 *iff* $\Pr(O|H1) > \Pr(O|H2)$.

With this in mind, let us proceed to examples and thought experiments.

Consider Modus Tollens:

If H, then O.

Not O

Therefore, not H.

Modus Tollens is a form of deductive inference. The first premise in an instance of this inference says that H is sufficient for O, and implies that O is necessary for H. Many forms of deductive inference similarly involve or imply statements of necessary conditions, sufficient conditions, or both. Popper's idea of falsifiability centers on Modus Tollens, using that inference to reject a hypothesis when a test fails to produce an observation entailed by the hypothesis. According to Popper, a hypothesis is falsifiable if and only if it is refutable by some conceivable observation. In contrast to Modus Tollens, consider another form of deductive inference, Modus Ponens. It says:

If O, then H.

O._____

H.

The first premise in Modus Ponens says O is sufficient for H. But in Popper's view, science is strictly *prohibitive*; a hypothesis can be falsified given a genuine counter-instance, but it is exceedingly rare for a hypothesis to be logically verified in virtue of Modus Ponens inference. This is because Modus Ponens involves stating that an observation deductively entails some hypothesis and yet observations very rarely entail the hypotheses that scientists test. Observations are typically deductively compatible with alternative hypotheses as well. As for Modus Tollens, it is now well-understood that scientific theories are rarely falsified on the basis of a single observation.³⁵ Just as observations rarely entail a hypothesis, a hypothesis rarely entails certain observations or the negation of them; in other words, the first premise in scientific applications of Modus Tollens would typically be false. So disconfirmation (and confirmation) in the context of scientific theories is better understood in probabilistic terms. Hypotheses make

certain observations more or less probable, rather than entailing them or their negations. This raises the question: can we turn to probabilistic versions of Modus Tollens and Modus Ponens?

The Lottery Paradox³⁶ shows us that we ought to be wary of using probabilistic versions of Modus Ponens in testing a hypothesis.³⁷ But interestingly, Probabilistic Modus Ponens has a close cousin. Here I express it as the updating rule within Bayesianism:

(Update Rule)

$\text{Pr}_{\text{time 1}}(H | O)$ is very high

O

O is all the evidence we have gathered between time 1 and time 2.

$\text{Pr}_{\text{time 2}}(H)$ is very high.

But some authors have endorsed the probabilistic version of Modus Tollens as a method of evidential reasoning. Prob-MT says:

$\text{Pr}(O | H)$ is very high

Not O

Not H.

Prob-MT requires the setting of a probability *cutoff* for the rejection of a hypothesis, and presumably, a justification for it. For instance, Richard Dawkins argues the probabilistic cutoff for theories pertaining to the origin of life on earth involves the number, n , of planets in the universe that are “suitable” for life.³⁸ A theory which posits the probability below $\frac{1}{n}$, Dawkins argues, should be rejected. The creationist Henry Morris proposes we assign theories which posit the number of times elementary particles in the universe could have changed state, a probability cutoff of $\frac{1}{10^{110}}$.³⁹ Meanwhile intelligent design theorist William Dembski assigns those same theories that Morris addresses with a probability cutoff of $\frac{1}{10^{150}}$.⁴⁰

But regardless of choice of cut-off, Prob-MT is an invalid form of inference. The invalidity stems from the conjunction of the probabilities of different observations over time.⁴¹ If H confers a very high probability (though less than unity) on each of the observations O1, O2, ..., O1000, it will confer a very low probability on their conjunction, assuming the observations are independent of each other conditional on H. Applying Prob-MT to scientific theories would render them with a very low probability after much testing, and eliminate virtually all well tested theories from science.⁴²

An underlying lesson from these problems with different types of inference is that support for a hypothesis is better understood *contrastively*.⁴³ To see why, consider the following thought experiment:

Suppose I send my valet to bring my urn containing 100 balls, of which only two are white. I draw one ball and find that it is white. Is this evidence against the hypothesis that he has brought the correct urn? And is $p = 0.02$ a proper measure of the strength of this evidence? Suppose that I keep in my urn vault two urns, one with two white balls and another, identical in appearance, that contains no white balls. Now is my observation of a white ball evidence that he has not brought the right urn? Fisher's disjunction still applies – either a rare event has occurred or the null hypothesis (correct) is false. But although the observation of a white ball is rare under the null hypothesis, it is even rarer under the alternative (wrong urn). In this case, the observation is actually strong evidence *in favour* of the null hypothesis.⁴⁴

The thought experiment illuminates an important idea about hypothesis testing: the better evidential reasoning approach to hypothesis testing tests a hypothesis *against another hypothesis*.

Testing is essentially a comparative process in which one tries to find observations that favour one hypothesis over another.⁴⁵

Prob-MT fails because it is not a comparative approach. And in addition to being an invalid form of reasoning under the conditions stated above, Prob-MT suffers from multiple counts of arbitrariness or *subjectivism*. To examine this subjectivism, I begin with some terminology. Prob-MT is a form of significance testing. Significance testing was proposed by R. A. Fisher⁴⁶ as a corrective measure for issues he saw in the Neyman-Person theory of hypothesis testing (discussion below). A hypothesis tested using significance testing is called the null hypothesis. This testing utilizes two values, the p -value and α value. The p -value is defined in terms of the *actual* observation produced by a test and other observations that were *possible* but didn't come about. But the p -value doesn't concern *all* possible observations. Rather it is the probability the test had of producing the actual observation *or* any of the other possible observations that were at least as improbable as the actual observation, if the null hypothesis is true. So it is the probability of a disjunction of actual and possible observations, conditional on the null hypothesis. If the hypothesis is true, what was the probability of attaining either the actual observation or any other that was at least as improbable? The answer equals the p -value.

The α value marks a conventionally chosen cut-off point, which is applied to the p -value. There are two interpretations of this. One interpretation is a yes-or-no affair; it says that if the p -value of the actual test outcome falls somewhere below the α value, then the null hypothesis should be rejected. The second interpretation introduces a matter of degree; it says that the lower the p -value is, the stronger the evidence against it, and that the α value marks the line between evidence against the null hypothesis being strong vs. not strong.

Both interpretations suffer from multiple counts of arbitrariness. Choosing a value for α is generally conceded as an arbitrary matter of convention. But how the outcome space for a significance test is carved up is also arbitrary, and some ways of carving in a given case can lead to the verdict that we reject (or have strong evidence against) the null, while other ways of carving oppositely imply that we do not reject (or do not have strong evidence against) the null. Howson and Urbach⁴⁷ provide the example of testing the null hypothesis that a coin is fair by flipping it 20 times. One way to carve up the space of possible outcomes is to say that getting 0 heads in 20 tosses is one possible outcome, getting 1 head in 20 tosses is another possible outcome, and so on. A different way to carve up the space is to say that getting 0 *or* 1 heads is one possible outcome, getting 1 *or* 2 heads is another possible outcome, and so on. Howson and Urbach show that on one such specification of the outcome space, and where α is set at 0.05, an actual outcome of 6 heads in 20 tosses would imply *rejecting* the hypothesis that the coin is fair, while on a different specification this same actual outcome and same value for α would imply *not* rejecting the hypothesis. The prospects for avoiding arbitrariness when opting for one carving rather than another, including prospects that appeal to some notion of the “most natural carving”, are bleak.⁴⁸

This example also shows that descriptions of the actual and possible data can come in varying degrees of logical strength, which relates to an additional problem. If significance testers insist on the logically strongest interpretation of data in each case, this can artificially result in p -values of unity, which would artificially (and thus illegitimately) protect a null hypothesis from ever being rejected by the chosen test. The problem is then a dilemma. On the one hand, the widely accepted principle of total evidence implies that any interpretation of a test should take account of everything we know, which implies using the logically strongest description of the

data; but we have just seen that if significance testers use such a description, they can entail illegitimate results. On the other hand, if they instead avoid such illegitimacy by sometimes opting against the logically strongest description of the data, then they must sometimes violate the principle of total evidence. The structure of significance testing ensures that this testing cannot always legitimately follow that principle.⁴⁹ Bayesianism, and more importantly, Likelihoodism, *do* conform with the principle of total evidence, and without entailing any illegitimacies.⁵⁰

Hereafter I presume that comparing Likelihoodism to its rivals by way of examples provides an adequate defense for believing Likelihoodism is true. The reasons for believing the normative advice stated by Likelihoodism typically consist in reasons for thinking applications of the Likelihood Principle produce more objective epistemic verdicts than produced by applying the tools of Likelihoodism's rivals. These reasons for believing Likelihoodism are thereby also reasons for believing some applications of its Likelihood Principle are objective – that they satisfy my objective condition.

Section 2.3.1: Inductive Risk and Royall's Three Questions

There are other important (albeit less technical) considerations involving subjective cutoffs in science concerning versions of the argument from “inductive risk”. The term “inductive risk” was coined by Hempel⁵¹ and the argument is most commonly attributed to Richard Rudner.⁵² Heather Douglas has recently developed her own version⁵³ and related arguments on what she calls the *direct* and *indirect* roles of values in science.⁵⁴ Simply put, the argument from inductive risk starts by noting that because science does not produce theories that are guaranteed to be absolutely true, there will always be a risk of error. Given the social aim not to inflict harm upon people, if some scientific theories produce evidence that is used to decide

whether some (for example) potentially hazardous chemical is safe, then there is reason to hold evidence about potentially hazardous substances to a higher standard of evaluation. But we avoid hazards because of our values, so this line of reasoning suggests that we determine thresholds for “sufficient evidence” by appeal to values.

In discussing that argument, authors of the science and values literature often draw a distinction between concepts of *belief* and *acceptance*.⁵⁵ Bayesianism, as we have already seen, says how you should change your *degrees of belief* as new observations come in. But there is a third concept that we can invoke when reasoning evidentially: *evidential favouring*. We have also discussed how *favouring* is an integral part to Likelihoodism. Statistician Richard Royall helpfully distinguishes between the concepts of acceptance, belief, and the concept of evidential favouring. In his 1997 book *Statistical Evidence: A Likelihood Paradigm*, Royall poses three questions concerning what scientists may want to address when evaluating observations:

1. What should you do?
2. What should you believe?
3. What do the observations tell you about the hypotheses you’re considering?

Question (1) falls in the domain of decision theory in which utilities (or non-epistemic considerations) as well as probabilities need to be considered, and relates most strongly to the arguments from inductive risk. Question (2), Royall thinks, is best suited for Bayesianism when certain conditions are met (see the discussion above on Bayesianism). And question (3) falls in the province of Likelihoodism and its Likelihood Principle. Through appealing only to evidential considerations, Likelihoodism does not tell us what we should believe (or whether we should accept a theory). Instead, it merely tells us what the evidence says and precludes using the candidate definitions of degree of confirmation, as previously noted.⁵⁶

Parsing things in this way helps clarify the modesty of the objectivity being proposed. Some applications of the Likelihood Principle that satisfy my objective condition will be restricted to a part of scientific practice that is different from those parts dealing with inductive risk, for instance. But distinguishing questions that scientists may pose also helps focus the debate on the descriptive argument I am making: if indeed the sorts of arguments for Likelihoodism (versus its competitors) that I have summarized *give us reasons* for believing Likelihoodism, as I have suggested, then – even if Likelihoodism is false without our knowing it – we thereby also have reasons for believing that applications of the Likelihood Principle can be objective in the sense I have identified. We have reason to interpret “favours” within the Likelihoodist framework as “objectively favours”.

Section 3.1: A Case Study

This section interprets a case-study using the Likelihood Principle, in order to illustrate how applications of it can be objective as I have argued. The case study involves two research approaches studying human aggressive behavior: a molecular genetics (MG) approach and a genetic/environmental (GE) approach.

Section 3.2: Applying the Likelihood Principle

The MG and the GE approaches compete to explain the primary cause of human aggression. More specifically, both research approaches attempt to answer the question of whether low brain activity of the enzyme monoamine oxidase A (MAOA) is a primary cause of trait aggression in human behavior. Trait aggression is defined as “an enduring disposition toward physical assault”.⁵⁷ Research in human behavior has focused on the enzyme MAOA and two MAOA genotypes because a primary function of the MAOA enzyme is to metabolize types of neurotransmitters in the brain and other organs.

Now to present the MG research approach and the hypotheses that it aims to test. One study conducted by the MG approach focuses on the men of a family with a rare genetic mutation, resulting in deletion in the gene encoding for MAOA. The researchers - Brunner and colleagues - conducting the study observed that the brothers of the family with the rare genetic mutation exhibited mental deficiencies and an inability to regulate impulsive aggressive behavior, while females within the family with the mutation had normal intelligence and normal behavior.⁵⁸ The hypothesis H1, that the MG approach is testing is:

H1: In most adult men with trait aggression, low MAOA genotype is the primary cause of their aggressive behavior.

In contrast, proponents of the GE approach endorse the following competing hypothesis:

H2: In most adult men with trait aggression, low MAOA genotype is **not** the primary cause of their aggressive behavior.

They endorse this competing hypothesis because they interpret their studies as giving them reason to believe that the MAOA genotype's causal influence is dependent on early childhood exposures to violence, and especially experience of child abuse. This is something to which H1 from the MG approach is not sensitive; that hypothesis presupposes that low MAOA genotype is a primary cause of the aggressive behavior even in the absence of early childhood exposure to violence.

This clarifies in a GE study focused on a birth cohort of 1,037 children (52% male) assessed at ages 3, 5, 7, 9, 11, 13, 15, 18 and 21.⁵⁹ Subjects are grouped according to low and high MAOA genotypes. Those two groups are each sub-divided into “No Childhood Maltreatment” (64%), “Probable Maltreatment” (28%) and “Severe Maltreatment” (8%). Data is sourced using methods appropriate to each stage of development (e.g., clinical diagnoses, personality checklists, individuals nominated by each subject that “...knows you well”, DSM-IV assessments, official convictions records). In each of the six participant sub-groups, four dependent variables are measured or assessed: (A) Conduct Disorder, (B) Disposition Towards Violence, (C) Anti-Social Personality Disorder symptoms (z scores) and (D) Convicted for Violent Offence.

Within this study, the competing hypotheses make competing predictions:

H1 Prediction: Men with low MAOA genotype and no childhood maltreatment, and men with low MAOA genotype and severe childhood maltreatment, will probably score **similarly** – both relatively high – in the categories of (A) Conduct Disorder and (B) Disposition Towards Violence.

H2 Prediction: Men with low MAOA genotype and no childhood maltreatment and men with low MAOA genotype and severe childhood maltreatment will probably score **differently** – low and high respectively – in the categories of (A) Conduct Disorder and (B) Disposition Towards Violence.

For each hypothesis to generate its prediction, it must be conjoined with various background assumptions, which I will collectively call ‘*Auxil1*’ for the MG approach and ‘*Auxil2*’ for the GE approach. In the actual experiments of the MG and GE approaches, these include estimates about potential experimental error and the effectiveness of controls. Additionally, the conclusions drawn from both tests which are causal and not merely correlational, include the assumption that the correlated factor is independent of the behavioral trait under investigation and that there is not a common cause of both.⁶⁰ At the very least, both approaches must assume that they allow other causal influences, but that other such influences are not among the primary ones.

Having generated the two conflicting predictions, the two hypotheses can be compared in terms of the following observation, gathered in the GE study:

O: Men with low MAOA genotype and no childhood maltreatment score **low** in (A) and (B), and men with low MAOA genotype and severe childhood maltreatment score **high** in (A) and (B).

This observation is graphically reported in the study’s Figure 2, reproduced here:

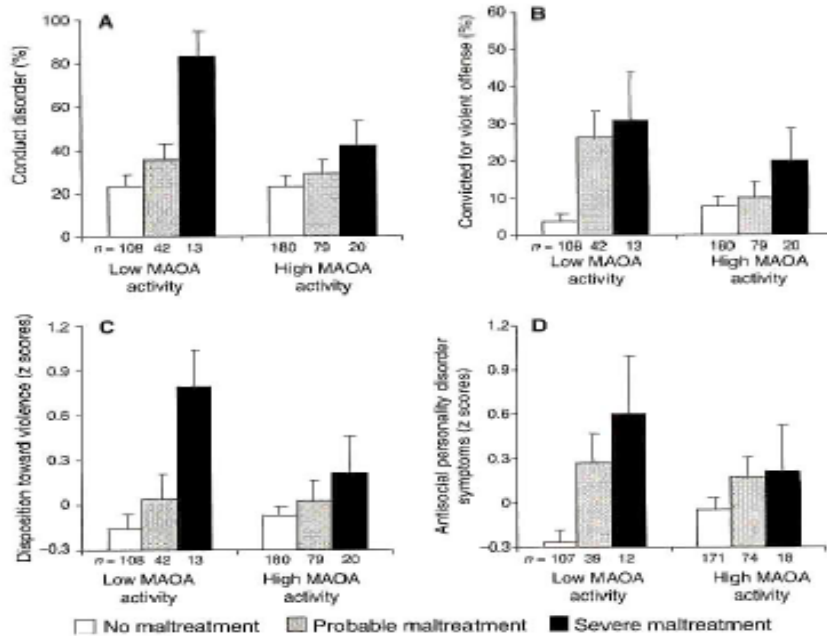


Fig. 2. The association between childhood maltreatment and subsequent antisocial behavior as a function of MAOA activity. (A) Percentage of males (and standard errors) meeting diagnostic criteria for Conduct Disorder between ages 10 and 18. In a hierarchical logistic regression model, the interaction between maltreatment and MAOA activity was in the predicted direction, $b = -0.63$, $SE = 0.33$, $z = 1.87$, $P = 0.06$. Probing the interaction within each genotype group showed that the effect of maltreatment was highly significant in the low-MAOA activity group ($b = 0.96$, $SE = 0.27$, $z = 3.55$, $P < 0.001$), and marginally significant in the high-MAOA group ($b = 0.34$, $SE = 0.20$, $z = 1.72$, $P = 0.09$). (B) Percentage of males convicted of a violent crime by age 26. The $G \times E$ interaction was in the predicted direction, $b = -0.83$, $SE = 0.42$, $z = 1.95$, $P = 0.05$. Probing the interaction, the effect of maltreatment was significant in the low-MAOA activity group ($b = 1.20$, $SE = 0.33$, $z = 3.65$, $P < 0.001$), but was not significant in the high-MAOA group ($b = 0.37$, $SE = 0.27$, $z = 1.38$, $P = 0.17$). (C) Mean z scores ($M = 0$, $SD = 1$) on the Disposition Toward Violence Scale at age 26. In a hierarchical ordinary least squares (OLS) regression model, the $G \times E$ interaction was in the predicted direction ($b = -0.24$, $SE = 0.15$, $t = 1.62$, $P = 0.10$); the effect of maltreatment was significant in the low-MAOA activity group ($b = 0.35$, $SE = 0.11$, $t = 3.09$, $P = 0.002$) but not in the high-MAOA group ($b = 0.12$, $SE = 0.07$, $t = 1.34$, $P = 0.17$). (D) Mean z scores ($M = 0$, $SD = 1$) on the Antisocial Personality Disorder symptom scale at age 26. The $G \times E$ interaction was in the predicted direction ($b = -0.31$, $SE = 0.15$, $t = 2.02$, $P = 0.04$); the effect of maltreatment was significant in the low-MAOA activity group ($b = 0.45$, $SE = 0.12$, $t = 3.83$, $P < 0.001$) but not in the high-MAOA group ($b = 0.14$, $SE = 0.09$, $t = 1.57$, $P = 0.12$).

Clearly, O bears out the $H2$ prediction much better than the $H1$ prediction. To put this in Likelihoodist terms, $H2+Auxil2$ conferred a higher probability on O than $H1+Auxil1$ did: $\Pr(O|H1+Auxil1) < \Pr(O|H2+Auxil2)$. Thus according to the Likelihood Principle, obtaining O then favoured $H2+Auxil2$ over $H1+Auxil1$.

Section 3.3: Objective Favouring

The case-study is an illustration of the application of the Likelihood Principle. My objectivism implies that the favouring of $H2$ over $H1$ by O is an objective favouring of exactly the sort specified in my objectivity condition from Section 2. Suppose that O from the case study accurately represents the world as intended, and suppose O is in fact more probable according to

$H2+Auxil2$ than to $H1+Auxil1$. Now imagine all researchers involved agreed with those two suppositions, but also all denied that O favours $H2+Auxil2$ over $H1+Auxil1$. (Perhaps their denial is motivated by how the advancement of egalitarian political goals would be better served by believing that the study results imply that O favours $H1+Auxil1$ over $H2+Auxil2$.) My position implies that this curious denial would be mistaken. Under the conditions I propose to be sufficient for objective favouring, O really would favour $H2+Auxil2$ over $H1+Auxil1$, despite the uniform denial of this. After all, within this particular testing contest, it is $H2+Auxil2$ that predicted O much more strongly.

Section 4.1: Objection

Given Longino's views summarized in Section 1.1 above, Longino would likely object to my argument by claiming the objectivity I propose is illusory. The favouring relations within the Likelihood framework depend on auxiliary assumptions that we could never *exhaustively* test or otherwise justify. Typically, an observation favours one hypothesis over another partly in virtue of the content of background assumptions that helped ensure the hypotheses lent different probabilities to the observation – predicted it to different degrees. Longino's view implies that any such favouring is objective only if we could independently justify *all* involved background assumptions, and she claims we cannot meet this demand. More formally, I will express Longino's objection as consisting in this *regress argument*:⁶¹

1. If:
 - a. applications of the Likelihood Principle depend on inferring predictions from pairs of compared hypotheses, and
 - b. the compared hypotheses typically must each be paired with many auxiliary assumptions in order to make predictions, and

c. independently justifying each and every auxiliary assumption serving the predictions would typically result in infinite regresses of justifications, then when application of the Likelihood Principle implies that an observation favours one hypothesis over another, this favouring is typically not objective.

2. Claim 1a is true.

3. Claim 1b is true.

4. Claim 1c is true.

5. When application of the Likelihood Principle implies that an observation favours one hypothesis over another, this favouring is typically not objective.

To illustrate this argument on Longino's behalf, let us consider how research approaches such as MG and GE can and often do draw upon complementary research programs in order to independently justify the background assumptions they use.

Such research approaches typically employ their own distinct set of auxiliary assumptions to help them arrive at their respective predictions. Some assumptions will be shared across competing research approaches, and some not. And while some are necessary for the predictions generated, others may not be necessary (e.g., different assumption could be subbed in for them, generating the same prediction) but are endorsed for one reason or another and do, in combination with other assumptions used, affect the estimates of the magnitudes of $\Pr(O|H1+Auxil1)$ and $\Pr(O|H2+Auxil2)$.

As briefly intimated above, one example of assumptions that research approaches employ are methodological assumptions. Research approaches studying human behaviour often employ assumptions about the degree of reliability of DNA extraction⁶² and sequencing⁶³ techniques, the validity of questionnaire manuals⁶⁴ and size and composition of their sample in relation to the target population. Usually, some methodological assumptions are highlighted by the authors as limitations of the study. And some methodological assumptions may be shared across research approaches while others may not.

An example of a set of methodological assumptions the GE approach employs that can affect the magnitudes of $\Pr(O|H1+Auxil1)$ and $\Pr(O|H2+Auxil2)$, and that are absent in the set of methodological assumptions employed by the MG approach, involve the conceptualization of indicators of environmental risk for developing aggressive behaviour. In the GE approach, conduct disorder is measured according to the DSM-IV. But changing the indicators of the environmental risk for developing conduct disorder can change the magnitudes of $\Pr(O|H1+Auxil1)$ and $\Pr(O|H2+Auxil2)$. This is illustrated by Foley et al. when they attempted to replicate the GE approach results while employing similar but distinct methodological assumptions. In their study, comprising 514 white male subjects aged 8 to 17, Foley et al. conducted interviews of the children and their parents on four occasions. On each occasion, indications of recent history (past 3 months) of conduct disorder are surveyed, and a history of exposure to environmental adversities and DNA are collected on occasions 3 and 4. Foley et al.'s research approach does not survey the variables that construct the maltreatment index applied by the GE approach. Instead, conduct disorder is measured in terms of parental neglect, exposure to interparental violence, and inconsistent parental discipline⁶⁵. Foley et al. did not find that

low MAOA correlates with conduct disorder, and this was regardless of childhood exposure to violence.⁶⁶

The results of other similar but distinct GE approaches are mixed,⁶⁷ highlighting potential issues of incommensurability that Longino cite,s given the absence of an agreed upon “best” set of methodological assumptions.

Additionally, there are assumptions about alternative causal influences that can affect the magnitudes of the probability values involved in predictions. For example, a child's MAOA genotype might be related to the chance of experiencing physical abuse indirectly via an association with parental characteristics that increase environmental risk exposure.⁶⁸ The GE prediction assumes children with low MAOA and severe childhood mistreatment exhibiting higher levels of reported aggressive behaviour do not disproportionately have parents who exhibit anti-social behaviour.

Fortunately, research approaches can and often do draw upon complementary research approaches in order to independently justify the background assumptions they use. One such approach performed by Kim-Cohen and colleagues, which I dub the Meta-Genetic Environmental approach (MGE), employs a meta-analysis.⁶⁹ Meta-analysis is a tool for determining whether a finding surmounts the limitations and differences across studies by pooling data to detect affects while avoiding overemphasis of estimates from any one study.⁷⁰ Meta-analysis continues to be a powerful method in the study of human behaviour.

Studies were included in the meta-analysis according to the criteria set by the MGE approach. The criteria are as follows: (1) the study must be published in a peer-reviewed journal; (2) the study must include genotypic information on the “variable number tandem repeat

polymorphism” (a type of DNA sequence) in the promoter region of the MAOA gene; (3) the study includes a measure of serious familial adversity in childhood that is significantly associated in a main effect fashion with the outcome measure; (4) the sample of each study is drawn from a non-clinical population. Included in the five studies that meet the meta-analysis criteria is the GE approach study.

Using standard methods to convert results of different studies into a common metric and to assess heterogeneity,⁷¹ the preliminary meta-analysis from the five studies conducted by the MGE approach finds supportive evidence of the GE approach’s results – an association between early familial adversity and mental health was significantly stronger in the low-activity MAOA vs the high-activity MAOA groups. This remained the case (to a slightly lesser degree) when the GE study was removed from the meta-analysis, and also when it and a study reporting even stronger association were both removed.

These meta-analyses suggest that across similar research approaches and their respective methodological assumptions (both dependently and independently of the GE approach), the association between maltreatment and aggressive behaviour is significantly stronger in the group of males with the genotype conferring low (rather than high) MAOA activity.

Researchers have also tested the assumption that children with low MAOA, experiences of severe childhood mistreatment, and who exhibit higher levels of aggressive behavior, also disproportionately have parents who exhibit anti-social behaviour. For example, Foley et al. found a significant correlation between child exposure to adversities and antisocial personality symptoms in the childrens’ mothers. This is consistent with their expectation that antisocial personality is associated with poor parenting. And adjusting for the main effects of the child’s MAOA genotype, the child’s level of exposure to adversity, and maternal antisocial personality

symptoms, did not lower the magnitude or the statistical significance of the association between conduct disorder and the interaction between MAOA activity and adversity.⁷²

These are just some of the great many examples of auxiliary assumptions being independently justified by neighbouring research programs. Longino is well aware of these practices. But her regress argument implies that in many cases of applying the Likelihood Principle, such practices could go on *ad infinitum*, and that this is a problem for an objectivist interpretation of applications of that principle.

Section 4.2: Response to the Longino Objection

To respond, I will argue against premise 1 in the regress argument. While I accept that justifications often could in principle go on *ad infinitum*, I deny that this undermines the type of objectivity I have endorsed.

There is a distinction between kinds of objectivity available in science.⁷³ One kind of objectivity is about the *content* involved in scientific reasoning - the particular contents of the concepts, theories, assumptions employed in science and their respective justifications. This is the kind of objectivity Longino's regress argument addresses. In effect, she is saying that just when you think you have objectively supported the content of one auxiliary assumption, there will be another you are depending upon and which in turn needs its contents justified in order for initial justification to objectively hold. But there is another kind of objectivity - one that I have argued for and which concerns the *form* of evidential reasoning according to the Likelihood Principle. If we assume O accurately represents the world as intended and O is in fact more probable according to H2+Auxil2 than to H1+Auxil1, then even if the *contents* of the auxiliary assumptions paired with the hypotheses are not (as Longino would have it) objectively supported, the three-place favouring *relation* that holds *between* O and H2+Auxil2 and

$H1+Auxil1$ would objectively hold. For this reason, Longino's likely objection would miss the mark.

Conclusion

I acknowledge there are important debates to be had about the many parts to science, and about values that often do and must play significant roles in those parts. For instance, it has long been widely acknowledged that values influence which research questions are asked in the first place, which hypotheses are tested, which tests are used, and so also which observations are gathered, and so on. When values influence such parts of scientific processes, they sometimes can change background assumptions and descriptions of observations. This can lead to changes in estimates of the magnitudes of $\Pr(O|H1+Auxil1)$ and $\Pr(O|H2+Auxil2)$ in particular cases. But those are issues about the estimations of the contents of particular applications of the Likelihood Principle, not about the *form* evidential favouring itself. I focus on the form of such favouring because this is at the heart of scientific reasoning.

Endnotes

¹ Throughout this paper, I will employ the distinction between epistemic and non-epistemic values. Assessments of the amount of support, likeliness or fruitfulness some theory or approach enjoys in terms of being true are considerations that count as “epistemic”. In contrast, ethical values, or the “acceptance” of a theory on the basis for a specific action are considerations that count as “non-epistemic” (see Ernan McMullin, “Values in Science,” *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association 1982* (1982): 3-28.). Note that some of the values that McMullin would consider to be epistemic, such as explanatory power and fertility, would be classified differently by other authors. See example, Heather E. Douglas, *Science, Policy, and the Value-Free Ideal* (Pittsburgh: University of Pittsburgh Press, 2009).

² Hans Reichenbach, *Experience and Prediction: An Analysis of the Foundations and the Structure of Knowledge* (United States: University of Chicago Press, 1938), 403. Reichenbach writes: “The mysticism of scientific discovery is nothing but a superstructure of images and wishes; the supporting structure below is determined by the inductive principle.”

Carl Hempel, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science* (New York: The Free Press, 1965), 91-92. Hempel famously argues the system of statements that represents scientific *knowledge* do not presuppose values, in the sense that values might provide evidential support for those statements. Rather, a scientific *method* might presuppose values, only in the sense that the rules of acceptance for deciding when a hypothesis has sufficient evidential support depend on assigning values to the outcomes that result from accepting or rejecting the hypothesis.

³ Douglas, *Science, Policy and the Value-Free Ideal*, 45-46. Douglas rejects the value free ideal for science, and the dominant position of the past forty years in philosophy of science “...the

notion that scientists are not involved in public life, that they provide no crucial advisory functions, and that they provide no assistance for decision making.”

⁴ Ibid., 45-46.; Matthew J. Brown, “Values in Science Beyond Underdetermination and Inductive Risk,” *Philosophy of Science* 80, no. 5 (2013): 836. Brown objects to Douglas who does not take the status of evidence as unproblematic, but posits any issues with the evidence are to be taken into account by prior consideration of values in selection of method and characterization of data. Brown further objects that this is based on “assum[ing] that in testing we ask given the evidence, what should we make of our hypothesis? Framed this way, values only play a role at the margins. This is a mistake since evidence can turn out to be bad in many ways: unreliable, noisy, laden with unsuitable concepts and interpretations, or irrelevant for the question at hand ... we can be skeptical about particular pieces or sets of evidence based on their clash with hypotheses, theories, or background assumptions that we have other good reasons to hold.” Also see Helen E. Longino, *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry* (Princeton: Princeton University Press), 1990 on intersubjective criticism.

⁵ See McMullin, “Values in Science,” 3-28.; Thomas S. Kuhn, “Objectivity, Value Judgment, and Theory Choice,” in *The Essential Tension: Selected Studies in Scientific Tradition and Change* (Chicago: University of Chicago Press, 1977), 320–39.

⁶ Longino, *Science as Social Knowledge*.; Peter K. Machamer and Gereon Wolters, *Science, Values, and Objectivity* (Pittsburgh: University of Pittsburgh Press, 2004).

⁷ Dave S. Collingridge and Edwin E. Gantt, “The Quality of Qualitative Research,” *American Journal of Medical Quality* 23, no. 5 (2008): 389-395.

⁸ Kevin C. Elliott, “Direct and Indirect Roles for Values in Science,” *Philosophy of Science* 78, no. 2 (2011): 303-324.

⁹ Richard Rudner, “The Scientist Qua Scientist Makes Value Judgments,” *Philosophy of Science* 20, no. 1 (1953): 1-6. One strategy is to revamp Richard Rudner’s Inductive Risk argument.

Building upon Rudner’s argument, Heather Douglas goes further to say there are two different roles (namely, direct and indirect) values can play in scientific reasoning. According to Douglas, values play a direct role when they act as “reasons in themselves to accept a claim” (see Douglas, *Science, Policy and the Value-Free Ideal*, 96). She emphasizes that values are relevant to formulating appropriate rules for accepting hypotheses in the face of inductive risk (i.e., in the face of the possibility that we might accept or reject hypotheses erroneously). The indirect role for values is relevant to numerous aspects of scientific activity, such as characterizing evidence and interpreting results, while the direct role of values include the ethical prohibitions against particular sorts of experimentation on humans and cases in which societal values discourage the development of new technologies (see Heather E. Douglas, “Inductive Risk and Values in Science,” *Philosophy of Science* 67, no. 4 (2000): 564).

Another approach is question the cogency of the fact/value distinction that Hempel famously argues for. For instance, Julian Dodd and Suzanne Stern-Gillet, “The Is/Ought Gap, the Fact/Value Distinction and the Naturalistic Fallacy,” *Dialogue: Canadian Philosophical Review/Revue Canadienne De Philosophie* 34, no. 4 (1995): 727-746. reject the fact/value distinction and argue that there can, at least in principle, be logical relationships between statements about “facts” and statements about “values” – see also Hillary Putnam (2004). In a recent response, Kevin C. Elliott and David Willmes, “Cognitive Attitudes and Values in Science,” *Philosophy of Science* 80, no. 5 (2013): 807-817. argue that despite perhaps the distinction being less rigid than Hempel thought, the sorts of non-epistemic values that typically influence science do not seem to provide evidential support for the sorts of claims that scientists

typically make. Elliott and Willmes give the example of a biological hypothesis not likely to be taken as true by scientists just because it advances egalitarian political goals. Helen E. Longino, "Gender, Politics, and the Theoretical Virtues," *Synthese* 104, no. 3 (1995): 383-397. questions the very attempt to formulate a convincing distinction between epistemic and non-epistemic values. See also Phyllis Rooney, "On Values in Science: Is the Epistemic/Non-Epistemic Distinction Useful?," in *Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association, Volume 2*, ed. David Hull, Micky Forbes and Kathleen Okruhlik (East Lansing: Philosophy of Science Association, 1992), 13-22.

Philosophers generally treat epistemic values such as consistency, simplicity, explanatory power and refutability as truth indicators (see McMullin, "Values in Science," 3-28.). But if the degree to which scientific values are weighed in a research project is the product of the type of problem being addressed and of the aims of the research project itself, then non-traditional values such as social utility, ontological heterogeneity, and novelty are no less epistemic (see Helen E. Longino, "Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy," In *Feminism, Science, and the Philosophy of Science*, ed. Lynn Hankinson Nelson and Jack Nelson (Dordrecht: Kluwer Academic Publishers, 1996), 39-58.).

¹⁰ Helen E. Longino, *The Fate of Knowledge*, (Princeton: Princeton University Press, 2002).; Miriam Solomon, *Social Empiricism*, (Cambridge, MA: MIT Press, 2001).

¹¹ Don Howard, "Lost wanderers in the forest of knowledge: Some thoughts on the discovery-justification distinction," *Revisiting discovery and justification* (2006): 3-22.; Longino, *Science as Social Knowledge*.; Longino, "Gender, Politics."; Longino, "Cognitive and Non-Cognitive Values," 39-58.; Longino, *The Fate of Knowledge*.; Helen E. Longino, "Lessons from Teaching

The Structure of Scientific Revolutions,” *Historical Studies in the Natural Sciences* 42, no. 5 (2012): 542-544.

¹² Richard Royall, “The strength of statistical evidence,” *The 52nd Session of the International Statistical Institute, Conference Proceedings 1999* (1999).

¹³ For an influential discussion of this, see Pierre Duhem, [1914] 1954, *The Aim and Structure of Physical Theory*, trans. from 2nd ed. by P. W. Wiener (Princeton: Princeton University Press, 1954), 187.

¹⁴ Kyle Stanford, “Underdetermination of Scientific Theory,” *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition), ed. Edward N. Zalta.

¹⁵ Longino, *Science as Social Knowledge*, 61.

¹⁶ *Ibid.*, 71.

¹⁷ Longino, *Fate of Knowledge*, 135.

¹⁸ *Ibid.*, 137.

¹⁹ *Ibid.*, 173-74.

²⁰ *Ibid.*, 207-08

²¹ Helen E. Longino, *Studying Human Behavior: How Scientists Investigate Aggression and Sexuality* (University of Chicago Press, 2013), 126.

²² *Ibid.*, 134-35.

²³ *Ibid.*, 119

²⁴ *Ibid.*, 118

²⁵ *Ibid.*, 21

²⁶ This argument strategy was recommended to me by Matthew Barker.

²⁷ I use the popular notation on which $\Pr(O|H1)$ means ‘the probability of O assuming H1 is true’. This is about how strongly the hypothesis predicts O, which is logically independent of how probably true the hypothesis is.

²⁸ Elliott Sober, *Evidence and Evolution: The Logic Behind the Science* (Cambridge: Cambridge University Press, 2008).

²⁹ Ian Hacking, *The Logic of Statistical Inference* (Cambridge: Cambridge University Press, 1965).; Anthony Edwards, *Likelihood* (Cambridge: Cambridge University Press, 1972).; Sober, *Evidence and Evolution: The Logic Behind the Science*.; Matthew Kotzen, “Selection biases in likelihood arguments,” *The British Journal for the Philosophy of Science* 63, no. 4 (2012): 825-839.; Benjamin C. Jantzen, “Piecewise versus Total Support: How to Deal with Background Information in Likelihood Arguments,” *Philosophy of Science* 81, no. 3 (2014): 313-331.

Whether the Likelihood Principle itself is something that can be true or false is controversial. I owe Ingo Brigandt for his specification that because the Likelihood Principle is a biconditional – an explicative definition – it may not be the kind of thing that can be true or false (they may only instead be truth-preserving).

³⁰ Malcolm Forster and Elliott Sober, “How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions,” *British Journal for the Philosophy of Science* 45 (1994): 1-36.

³¹ Sober, *Evidence and Evolution: The Logic Behind Science*.

³² Ibid., 31. Objective Bayesians think the form of Bayesian reasoning is objectively correct - how your beliefs should change as new evidence rolls in. It is nonetheless consistent that the content may sometimes consist of subjective components (e.g., subjective prior probabilities).

³³ Ibid., 26.

³⁴ By “reliable”, I mean just that the base rate has been accounted for and the number of tests administered is representative of the population.

³⁵ D. Gilles, “The Duhem Thesis and the Quine Thesis,” in *Philosophy of Science in the Twentieth Century* (Oxford: Blackwell Publishers, 1993). 98–116.

³⁶ (Prob-MP) says: $\Pr(H | O)$ is very high; O ; therefore, $\Pr(H)$ is very high. While Prob-MP is valid, the Lottery Paradox gives reason to be cautious. Suppose there are 1,000 lottery tickets sold in a lottery. Suppose also that the lottery is fair and you believe it is fair (you believe each ticket has an equal chance of winning). If high probability is justification for belief, then you are justified in believing that no. 1 ticket will not win because the probability of no. 1 ticket winning is $\frac{1}{1000}$. You are also justified in believing that no. 2 ticket will not win because no. 2 ticket has the same probability as no. 1 ticket and so on. But if you conjoin the 1000 beliefs that pertain to each lottery ticket (the belief that the ticket will not win) together with your other beliefs (the lottery is fair), then your beliefs are contradictory: you believe that some ticket will win (given your belief that the lottery is fair) and you believe no ticket will win (Henry E. Kyburg, “Conjunctivitis,” in *Induction, Acceptance, and Rational Belief*, ed. M. Swain (Dordrecht: Reidel, 1970): 55-82.; see also Sober, *Evidence and Evolution: The Logic Behind Science*).

³⁷ Another reason to be cautious of Prob-MP is that it is possible to have a $\Pr(H)$ conditional on O be very high while also having $\Pr(O)$ unconditionally be very low. For example, “[e]ven though it is very probable that the roulette wheel ball landed double-zero on the last spin, given that your honest and visually acute friend told you that this is what happened, it is still unconditionally improbable that the ball landed double-zero” (Elliott Sober, “Intelligent design and probability reasoning,” *International Journal for Philosophy of Religion* 52, no.2 (2002): 60.).

-
- ³⁸ Richard Dawkins, *The Blind Watchmaker* (New York: Norton, 1986). 144-6.
- ³⁹ Henry M. Morris, *King of Creation* (San Diego, Calif.: CLP Publishers, 1990).
- ⁴⁰ Willian A. Dembski, *The Design Inference* (Cambridge: Cambridge University Press, 1998).
- ⁴¹ Hacking, *The Logic of Statistical Inference.*; Edwards, *Likelihood.*; Royall, *Statistical Evidence: A Likelihood Paradigm*, Vol. 71, CRC press, 1997.
- ⁴² *Ibid.*, 51.
- ⁴³ *Ibid.*, 33.
- ⁴⁴ Royall, *Statistical Evidence: A Likelihood Paradigm*, 67.
- ⁴⁵ Elliot Sober, "Contrastive Empiricism," In *Minnesota Studies in the Philosophy of Science: Scientific Theories*, ed. W. Savage (Minneapolis: University of Minnesota Press, 1990): 392-412.; Elliot Sober, "Modus Darwin," *Biology and Philosophy* 14, no. 2 (1999): 253–78.; Elliot Sober, "Two Uses of Unification," in *The Vienna Circle and Logical Empiricism, Vienna Circle Institute Yearbook, vol 10*, ed. F. Stadler (Dordrecht: Springer, 2003), 205-216.
- ⁴⁶ Ronald. A Fisher, *Statistical methods and scientific inference*.
- ⁴⁷ Colin Howson and Peter Urbach, *Scientific reasoning: the Bayesian approach*, 136-7.
- ⁴⁸ Sober, *Evidence and Evolution: The Logic Behind the Science*, 55.
- ⁴⁹ I owe the specification of the dilemma to Matthew Barker.
- ⁵⁰ Sober, *Evidence and Evolution: The Logic Behind the Science*, 56
- ⁵¹ Hempel, *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*.
- ⁵² Rudner, "The Scientist Qua Scientist Makes Value Judgments," 1-6.
- ⁵³ Douglas, "Inductive Risk and Values in Science," 559-579.
- ⁵⁴ Douglas, *Science, Policy, and the Value-Free Ideal*.

⁵⁵ Kevin Elliott and David Wilmes (2013) propose a descriptive account of acceptance whereby S accepts that *h*, if S presupposes *h* for specific reasons in her deliberations. Elliott and Wilmes claim that two main differences between belief and acceptance follow from this account. First, belief cannot be controlled at will whereas acceptance of a proposition can be voluntary. Second, belief varies on the amount of evidence available and is not based on other contextual factors because belief is always focused on arriving at truth (Bernard Williams, “Deciding to Believe,” in *Problems of the Self*, ed. Bernard Williams (Cambridge: Cambridge University Press, 1973), 136-151.; Michael Bratman, *Faces of Intention* (Cambridge: Cambridge University Press, 1999). Daniel Steel, “Acceptance, values, and inductive risk,” *Philosophy of Science* 80, no. 5 (2013): 818-828. advocates for a similar distinction and adopts a theory of acceptance developed by Jonathan Cohen. Steel argues that non-epistemic values can legitimately influence scientific inferences because one central aim of scientific inference is to decide whether to accept or reject hypotheses. And decisions about whether to accept or reject a scientific hypothesis can have implications for practical action. When this happens, acceptance decisions should depend in part on non-epistemic value judgements about the cost of error.

⁵⁶ Sober, *Evidence and Evolution: The Logic Behind Science*.

⁵⁷ Caspi et al., “Role of genotype in the cycle of violence in maltreated children,” *Science* 297, no. 5582 (2002): 851-854.; Moffitt et al., “Males on the life-course-persistent and adolescence-limited antisocial pathways: Follow-up at age 26 years,” *Development and psychopathology* 14, no. 1 (2002): 179-207.

⁵⁸ Brunner et al., “Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A,” *American Association for the Advancement of Science* 262, no. 5133 (1993): 578-580.

⁵⁹ Caspi et al., “Role of genotype in the cycle of violence in maltreated children.”

⁶⁰ Longino, “Lessons from Teaching The Structure of Scientific Revolutions,” 101.

⁶¹ I owe this specification of Longino’s likely objection to Matthew Barker.

⁶² B. Freeman et al., “DNA from Buccal Swabs Recruited by Mail: Evaluation of Storage Effects on Long-Term Stability and Suitability for Multiplex Polymerase Chain Reaction Genotyping,” *Behavior Genetics* 33, no. 1 (2003): 67-72.

⁶³ Sue. Z. Sabol, Stella Hu, and Dean Hamer, “A Functional Polymorphism in the Monoamine Oxidase A Gene Promoter,” *Human Genetics* 103, no. 3 (1998): 273-279.

⁶⁴ Sara R. Jaffee et al., “Physical Maltreatment Victim to Antisocial Child: Evidence of an Environmentally Mediated Process,” *Journal of Abnormal Psychology* 113, no. 1 (2004): 44-55.

⁶⁵ Foley et al., “Childhood Adversity, Monoamine Oxidase A Genotype, and Risk for Conduct Disorder,” 739. “Interparental violence, and inconsistent parental discipline. All 3 adversities were assessed at wave 3 of the VTSABD with the same assessment method (personal interview) and time frame of survey (ever). Three items rated by parents were used to define parental neglect: (1) Did anyone ever say you weren’t looking after the children properly? (2) Has anyone ever thought that 1 of the children became ill because the child wasn’t looked after properly or because the home wasn’t clean enough? (3) Was there ever a time when 1 of the children was very ill, but, at the time, you didn’t think the child needed to see a doctor, or was there?”

⁶⁶Ibid., 741-2.

⁶⁷ Brett C. Haberstick et al., “Monoamine Oxidase A (MAOA) and Antisocial Behaviors in the Presence of Childhood and Adolescent Maltreatment,” *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 135B, no. 1 (2005): 59-64.; Kent W. Nilsson et al., “Role of Monoamine Oxidase A Genotype and Psychosocial Factors in Male Adolescent

Criminal Activity,” *Biological Psychiatry* 59, no. 2 (2006): 121-127.; Susan E. Young et al., “Interaction between MAO-A Genotype and Maltreatment in the Risk for Conduct Disorder: Failure to Confirm in Adolescent Patients,” *American Journal of Psychiatry* 163, no. 6 (2006): 1019-1025.

⁶⁸ J. Kim-Cohen et al., “MAOA, Maltreatment, and Gene-Environment Interaction Predicting Children's Mental Health: New Evidence and a Meta-Analysis,” *Molecular Psychiatry* 11, no. 10 (2006): 903-913.

⁶⁹ Ibid.

⁷⁰ Ibid., 904.

⁷¹ Ibid., 909.

⁷² Foley et al., “Childhood Adversity, Monoamine Oxidase A Genotype, and Risk for Conduct Disorder”.

⁷³ This response to Longino’s regress argument was recommended by Matthew Barker.

Bibliography

Works Cited

- Barker, Matthew J. September 1, 2017, in correspondence, Montreal, Quebec, Concordia University.
- Bratman, Michael. *Faces of Intention*. Cambridge: Cambridge University Press, 1999.
- Brown, Matthew J. "Values in Science Beyond Underdetermination and Inductive Risk." *Philosophy of Science* 80, no. 5 (2013): 829-839.
- Brunner, Han G., M. Nelen, X. O. Breakefield, H. H. Ropers, and B. A. Van Oost. "Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A." *American Association for the Advancement of Science* 262, no. 5133 (1993): 578-580.
- Caspi, Avshalom, Joseph McClay, Terrie E. Moffitt, Jonathan Mill, Judy Martin, Ian W. Craig, Alan Taylor, and Richie Poulton. "Role of genotype in the cycle of violence in maltreated children." *Science* 297, no. 5582 (2002): 851-854.
- Collingridge, Dave S. and Edwin E. Gantt. "The Quality of Qualitative Research." *American Journal of Medical Quality* 23, no. 5 (2008): 389-395.
- Dawkins, Richard. *The Blind Watchmaker*. New York: Norton, 1986.
- Dembski, William A. *The Design Inference*. Cambridge: Cambridge University Press, 1998.
- Dodd, Julian and Suzanne Stern-Gillet. "The Is/Ought Gap, the Fact/Value Distinction and the Naturalistic Fallacy." *Dialogue: Canadian Philosophical Review/Revue Canadienne De Philosophie* 34, no. 4 (1995): 727-746.
- Douglas, Heather E. "Inductive Risk and Values in Science." *Philosophy of Science* 67, no. 4 (2000): 559-579.
- . *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press, 2009.
- Duhem, Pierre. *The Aim and Structure of Physical Theory*. Translated by P.P. Wiener. Princeton: Princeton University Press, 1954.
- Edwards, Anthony. *Likelihood*. Cambridge: Cambridge University Press, 1972.
- Elliott, Kevin C. "Direct and Indirect Roles for Values in Science." *Philosophy of Science* 78, no. 2 (2011): 303-324.

- Elliott, Kevin C. and David Willmes. "Cognitive Attitudes and Values in Science." *Philosophy of Science* 80, no. 5 (2013): 807-817.
- Fisher, Ronald A. *Statistical methods and scientific inference*. 1959.
- Foley, Debra L., Lindon J. Eaves, Brandon Wormley, Judy L. Silberg, Hermine H. Maes, Jonathan Kuhn, and Brien Riley. "Childhood Adversity, Monoamine Oxidase A Genotype, and Risk for Conduct Disorder." *Arch Gen Psychiatry* 61, no.7 (2004):738–744. doi:10.1001/archpsyc.61.7.738.
- Forster, Malcolm and Elliott Sober. "How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions." *British Journal for the Philosophy of Science* 45 (1994): 1-36.
- Freeman, B., N. Smith, C. Curtis, L. Hockett, J. Mill, and I. W. Craig. "DNA from Buccal Swabs Recruited by Mail: Evaluation of Storage Effects on Long-Term Stability and Suitability for Multiplex Polymerase Chain Reaction Genotyping." *Behavior Genetics* 33, no. 1 (2003): 67-72.
- Gilles, D. "The Duhem Thesis and the Quine Thesis." In *Philosophy of Science in the Twentieth Century*, 98–116. Oxford: Blackwell Publishers, 1993.
- Haberstick, Brett C., Jeffrey M. Lessem, Christian J. Hopfer, Andrew Smolen, Marissa A. Ehringer, David Timberlake, and John K. Hewitt. "Monoamine Oxidase A (MAOA) and Antisocial Behaviors in the Presence of Childhood and Adolescent Maltreatment." *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 135B, no. 1 (2005): 59-64.
- Hacking, Ian. *The Logic of Statistical Inference*. Cambridge: Cambridge University Press, 1965.
- Hempel, Carl. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: The Free Press, 1965.
- Howard, Don. "Lost wanderers in the forest of knowledge: Some thoughts on the discovery-justification distinction." *Revisiting discovery and justification* (2006): 3-22.
- Howson, Colin, and Peter Urbach. *Scientific reasoning: the Bayesian approach*. Open Court Publishing, 2006.
- Jaffee, Sara R., Avshalom Caspi, Terrie E. Moffitt, and Alan Taylor. "Physical Maltreatment Victim to Antisocial Child: Evidence of an Environmentally Mediated Process." *Journal of Abnormal Psychology* 113, no. 1 (2004): 44-55.
- Jantzen, Benjamin C. "Piecewise versus Total Support: How to Deal with Background Information in Likelihood Arguments." *Philosophy of Science* 81, no. 3 (2014): 313-331.

- Kim-Cohen, J., Avshalom Caspi, A. Taylor, B. Williams, R. Newcombe, I. W. Craig, and Terrie E. Moffitt. "MAOA, Maltreatment, and Gene-Environment Interaction Predicting Children's Mental Health: New Evidence and a Meta-Analysis." *Molecular Psychiatry* 11, no. 10 (2006): 903-913.
- Kotzen, Matthew. "Selection biases in likelihood arguments." *The British Journal for the Philosophy of Science* 63, no. 4 (2012): 825-839.
- Kuhn, Thomas S. "Objectivity, Value Judgment, and Theory Choice." In *The Essential Tension: Selected Studies in Scientific Tradition and Change*, 320–39. Chicago: University of Chicago Press, 1977.
- Kyburg, Henry E. "Conjunctivitis." In *Induction, Acceptance, and Rational Belief*, edited by M. Swain, 55-82. Dordrecht: Reidel, 1970.
- Longino, Helen E. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press, 1990.
- . "Gender, Politics, and the Theoretical Virtues." *Synthese* 104, no. 3 (1995): 383-397.
- . "Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy." In *Feminism, Science, and the Philosophy of Science*, 39-58. Edited by Lynn Hankinson Nelson and Jack Nelson. Dordrecht: Kluwer Academic Publishers, 1996.
- . *The Fate of Knowledge*. Princeton: Princeton University Press, 2002.
- . "Lessons from Teaching The Structure of Scientific Revolutions." *Historical Studies in the Natural Sciences* 42, no. 5 (2012): 542-544.
- . *Studying Human Behavior: How Scientists Investigate Aggression and Sexuality*. University of Chicago Press, 2013.
- Machamer, Peter K. and Gereon Wolters. *Science, Values, and Objectivity*. Pittsburgh: University of Pittsburgh Press, 2004.
- McMullin, Ernan. "Values in Science." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1982 (1982): 3-28.
<http://www.jstor.org/stable/192409>.
- Moffitt, Terrie E., Avshalom Caspi, Honalee Harrington, and Barry J. Milne. "Males on the life-course-persistent and adolescence-limited antisocial pathways: Follow-up at age 26 years." *Development and psychopathology* 14, no. 1 (2002): 179-207.
- Morris, Henry M. *King of Creation*. San Diego, Calif: CLP Publishers, 1990.

- Nilsson, Kent W., Rickard L. Sjöberg, Mattias Damberg, Jerzy Leppert, John Öhrvik, Per Olof Alm, Leif Lindström, and Lars Oreland. "Role of Monoamine Oxidase A Genotype and Psychosocial Factors in Male Adolescent Criminal Activity." *Biological Psychiatry* 59, no. 2 (2006): 121-127.
- Putnam, Hilary. *The Collapse of the fact/value Dichotomy and Other Essays*. 1st Harvard University Press pbk. ed. Harvard University Press, 2004.
- Reichenbach, Hans. *Experience and Prediction: An Analysis of the Foundations and the Structure of Knowledge*. United States: University of Chicago Press, 1938.
- Rooney, Phyllis. "On Values in Science: Is the Epistemic/Non-Epistemic Distinction Useful?" In *Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association, Volume 2*, 13-22. Edited by David Hull, Micky Forbes, and Kathleen Okruhlik. East Lansing: Philosophy of Science Association, 1992.
- Royall, Richard. "The strength of statistical evidence." *The 52nd Session of the International Statistical Institute, Conference Proceedings 1999* (1999).
<https://www.stat.fi/isi99/proceedings/arkisto/varasto/roya0578.pdf>
- . *Statistical Evidence: A Likelihood Paradigm*. Vol. 71. CRC press, 1997.
- Rudner, Richard. "The Scientist Qua Scientist Makes Value Judgments." *Philosophy of Science* 20, no. 1 (1953): 1-6.
- Sabol, Sue. Z., Stella Hu, and Dean Hamer. "A Functional Polymorphism in the Monoamine Oxidase A Gene Promoter." *Human Genetics* 103, no. 3 (1998): 273-279.
- Sober, Elliott. *Evidence and Evolution: The Logic Behind the Science*. Cambridge University Press, 2008.
- . "Two Uses of Unification." In *The Vienna Circle and Logical Empiricism, Vienna Circle Institute Yearbook, vol 10*, edited by F. Stadler, 205-216. Dordrecht: Springer, 2003.
- . "Intelligent design and probability reasoning." *International Journal for Philosophy of Religion* 52, no. 2 (2002): 65-80.
- . "Modus Darwin." *Biology and Philosophy* 14, no. 2 (1999): 253-78.
<https://doi.org/10.1023/A:1006626115852>.
- . "Contrastive Empiricism." In *Minnesota Studies in the Philosophy of Science: Scientific Theories*, edited by W. Savage, 392-412. Minneapolis: University of Minnesota Press, 1990.
- Solomon, Miriam. *Social Empiricism*. Cambridge, MA: MIT Press, 2001.

Stanford, Kyle. "Underdetermination of Scientific Theory." *The Stanford Encyclopedia of Philosophy* (Spring 2016 Edition). Edited by Edward N. Zalta.
<https://plato.stanford.edu/archives/win2017/entries/scientific-underdetermination>

Steel, Daniel. "Acceptance, values, and inductive risk." *Philosophy of Science* 80, no. 5 (2013): 818-828.

Williams, Bernard. "Deciding to Believe." In *Problems of the Self*, edited by Bernard Williams, 136-151. Cambridge: Cambridge University Press, 1973.

Young, Susan E., Thomas J. Crowley, Robin P. Corley, Michael C. Stallings, Brett C. Haberstick, Andrew Smolen, and John K. Hewitt. "Interaction between MAO-A Genotype and Maltreatment in the Risk for Conduct Disorder: Failure to Confirm in Adolescent Patients." *American Journal of Psychiatry* 163, no. 6 (2006): 1019-1025.